

Outbound Internet Access and Proxy

Date published: 2019-08-22

Date modified:



Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

Outbound internet access and proxy.....	4
Using a non-transparent proxy.....	4
Setting up a non-transparent proxy in CDP.....	5
Setting up a web proxy for TLS inspection.....	9

Outbound internet access and proxy

This section provides information on the outbound network destinations for CDP, and instructions on how to configure CDP to use a proxy for outbound access.

Depending on your enterprise requirements, you may have limited or restricted outbound network access and/or require the use of an internet proxy. Registering a cloud provider environment in CDP, as well as creating clusters within an environment requires outbound network access to certain destinations, and in some cases must go through a proxy.

Scenario	Documentation
My environment has limited outbound internet access	Refer to the following documentation for information on network rules: <ul style="list-style-type: none"> • AWS outbound network access destinations • Azure outbound network access destinations • GCP outbound network access destinations
My environment requires use of a proxy for outbound internet access	Refer to Using a non-transparent proxy on page 4.

Related Information

[Using a non-transparent proxy](#)

Using a non-transparent proxy

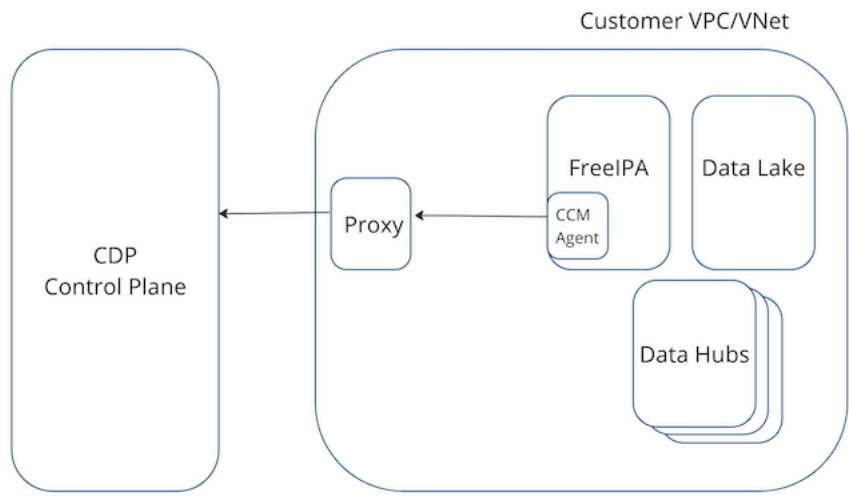
Refer to this section if your environment requires all internet traffic to go through an internet proxy. You can use a proxy server to control the connections that are allowed from your VPC or VNet and block unauthorized connections initiated from your environment.


When creating a CDP environment, you can set up an HTTP proxy such as Squid or a comparable product. For a majority of use cases, this is enough to direct the traffic through a proxy.

Proxy servers can be used for:

- FreeIPA backups: Backups created on an hourly basis are uploaded to cloud storage S3/ADLS Gen2.
- Parcel downloads: Although CDP currently only supports pre-warmed images, it is a requirement to download parcels from archive.cloudera.com when an upgrade is performed.
- Cluster Connectivity Manager (CCM): Communication via CCMv1 and CCMv2.
- TLS and Deep Packet Inspection (DPI): TLS and DPI Inspection can be performed through the use of a proxy. To see how to configure this, refer to the [Setting up a web proxy for TLS inspection](#) section below.

The following diagram illustrates the communication between the customer's CDP environment and the CDP Control Plane in a cloud provider network (VPC/VNet) via a web proxy:



 **Note:** The above diagram illustrates a simple architecture where the web proxy is located within the customer VPC/VNet. However, the web proxy could be located anywhere in the network between the CCM agent and the CCM server: in the customer VPC outside of CDP (as above), in another customer VPC, or in a cloud-based web proxy service.

Supported CDP services


The following CDP services allow the use of a web proxy:

CDP service	AWS	Azure	GCP
Data Lake	GA	GA	GA
FreeIPA	GA	GA	GA
Data Engineering	GA		
Data Hub	GA	GA	GA
Data Warehouse	GA		
DataFlow	GA		
Machine Learning	GA		
Operational Database			

Note that in order to use a non-transparent proxy with CDP data services (such as Data Engineering, Data Warehouse, DataFlow, and Machine Learning), you must first configure it at the environment level and then once again when enabling/activating the CDP data service.

Setting up a non-transparent proxy in CDP

To set up a proxy server you can register an http proxy server as a shared resource and then add that shared resource when you set up your environment.


 **Note:** Once you register a proxy in CDP, there is no option to edit the proxy registration. You need to delete the proxy registration in CDP and register the proxy again.

Required role: EnvironmentCreator can register a proxy in CDP and manage user access to the proxy. Owner or SharedResourceUser can view the proxy details. Owner can delete the proxy registration from CDP.

Steps

For CDP UI

1. Log in to the CDP web interface.
2. Navigate to the Management Console.
3. Select Shared Resources > Proxies from the left navigation pane.
4. Click Create Proxy Configuration.

 **CLOUDERA**
Management Console

Dashboard

Environments

Data Lakes

User Management

Data Hub Clusters

Classic Clusters

Audit

Consumption

Shared Resources

Cluster Templates

Cluster Template Overrides

Proxies

Credentials

Recipes

Image Catalogs

Global Settings

Notifications

Get Started

Help

2.91.0-b107

Proxies / Create

Create Proxy

Name *

Enter name

Description

Enter description

Protocol *

HTTP

Server Host *

Enter Server Host

Server Port *

Enter Server Port

No Proxy Hosts

Enter No Proxy Hosts

Inbound Proxy CIDR

Enter custom CIDR IP range to allow inbound communication

Username

Enter Username


Password

Enter Password

REGISTER

5. Enter the information for your proxy server:

Parameter	Description
Name (Required)	Provide a name for the proxy. The name will be used for this specific proxy in CDP.
Description	You can optionally specify a longer description for this proxy.

Parameter	Description
Protocol (Required)	Select the protocol used by the proxy: HTTP or HTTPS.
Server Host (Required)	Provide proxy server's host.
Server Port (Required)	Provide the proxy server's port.
No Proxy Hosts	<p>The no-proxy field allows you to designate specific IP addresses, domains, or subdomains that bypass the proxy. This setting can be useful for locally resolvable and internal endpoints, for example the CCMv2 agent or the metering agent.</p> <p>Enter the values for this field in a comma-separated list. For example: 172.100.0.110,domainname.com,my.host.com</p> <p>Note the following guidelines:</p> <ul style="list-style-type: none"> The period character (".") is allowed as a prefix for domain names only CIDR notation is not allowed <p> Note: If you are running a CDP environment on Azure using Runtime 7.2.14 or newer and you don't want to make an exception for management.azure.com (which is mentioned as required in Azure outbound network access destinations), you can add the following to the No Proxy Hosts list:</p> <pre>localhost,127.0.0.1,169.254.169.254,168.63.129.16</pre>
Inbound Proxy CIDR	Provide a custom CIDR IP range to allow inbound communication. Required when you use an FQDN instead of an IP address to define your proxy. Without providing the CIDR IP range security groups cannot be adjusted to allow communication with the proxy and Kubernetes server.
User name	If needed, provide a user name to access the proxy.
Password	If needed, provide a password to access the proxy.

6. Click REGISTER.
7. Click Environments in the left navigation pane, then click Register Environment.
8. Add your environment information, navigating through the Register Environment and Data Lake Scaling steps.

9. When you reach the Region, Networking and Security steps, choose the Proxy you registered.

The screenshot shows the 'Environments / Environments' page in the CDP console. On the left, a sidebar lists the steps: 'Register Environment' (completed), 'Data Lake Scaling' (completed), and 'Region, Networking, Security and Storage' (current step). The main content area is divided into three sections: 'Region, Location' with a 'Select Region' dropdown set to 'North Europe - North Europe'; 'Network' with a 'Select Network' dropdown set to 'Create new network', a 'Network CIDR' input field set to '10.10.0.0/16', and two toggle switches for 'Create Private Subnets' and 'Enable Cluster Connectivity Manager'; and 'Proxies' with a 'Select Proxy Configuration' dropdown set to 'Do not use Proxy Configuration'. A yellow arrow points to this dropdown.

10. Finish setting up your Environment.

For CDP CLI

As an alternative to using the UI, you can also register the proxy using the CLI.

1. Use the following commands:

```
cdp environments create-proxy-config \
  --proxy-config-name companyProxy \
  --host 10.102.0.19 \
  --port 3128 \
  --user squid \
  --password squid \
  --protocol http
```

2. Provide the proxyConfigName in the environment JSON:

```
{
  ...
  "subnetIds": [
    "subnet-1",
    "subnet-2",
    "subnet-3"
  ],
  "proxyConfigName": "companyProxy" must be on the root level
}
```

3. Or in the --proxy-config-name argument of the environment creation command, enter the following:

AWS:

```
cdp environments create-aws-environment \
  --cli-input-json '{...}' \
```



```
--proxy-config-name companyProxy
```

Azure:

```
cdp environments create-azure-environment \
  --cli-input-json '{...}' \
  --proxy-config-name companyProxy
```



Note: These proxy settings do not apply to cluster recipes. If you planning to use the recipes, then you can set the proxy settings manually, if needed. You can find the proxy settings in the `/etc/cdp/proxy.env` file.

Related Information

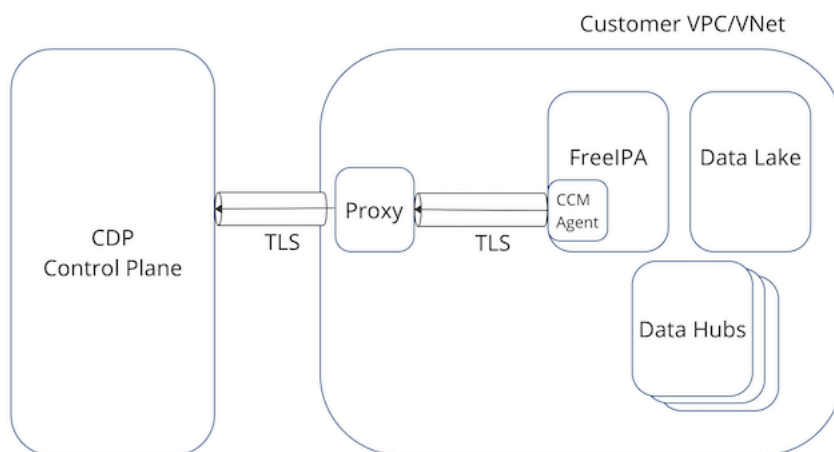
[Use a non-transparent proxy with CML on AWS environments](#)

Setting up a web proxy for TLS inspection

After setting up the proxy server in CDP, you can further configure it to perform TLS interception and Deep Packet Inspection (DPI).

Without a web proxy, a single TLS session is initiated from the CCM agent and terminated at the CCM server within the Control Plane. With the introduction of the web proxy, there are two TLS sessions: (1) a TLS session initiated from the CCM agent terminating at the proxy and (2) a TLS session initiated from the proxy terminating at the CCM server within the Control Plane. The web proxy decrypts the packets of the TLS session, performs any operations on the clear text (such as DPI), and re-encrypts the packets onto the second TLS session. Thus the proxy behaves as a man-in-the-middle (MITM) that is able to view the communications between the CCM agent and the CCM server using TLS inspection.

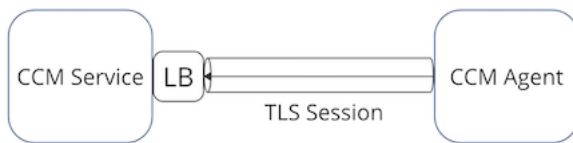
The following diagram illustrates the communication between the customer's CDP environment and the CDP Control Plane in a cloud provider network (VPC/VNet) via a web proxy:



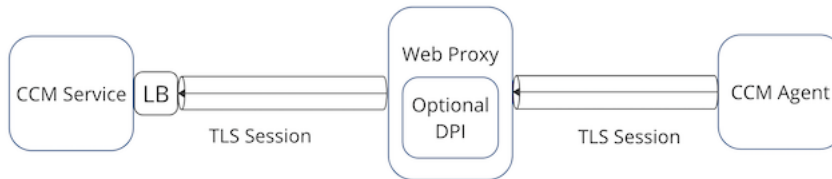
Note: The above diagram illustrates a simple architecture where the web proxy is located within the customer VPC/VNet. However, the web proxy could be located anywhere in the network between the CCM agent and the CCM server: in the customer VPC outside of CDP (as above), in another customer VPC, or as a cloud-based web proxy service.

The CDP architecture with and without proxy-based TLS inspection is illustrated in the following two diagrams.

The following diagram illustrates CCM communication without a web proxy as MITM:



The following diagram illustrates CCM communication with a web proxy as MITM:



To configure TLS inspection, you need to set up your proxy to trust the certificate of CCM, and, in turn, make sure that CCM trusts the proxy's CA certificate.

Steps

1. Register a new CDP environment.
2. After the FreeIPA nodes are running, SSH into the FreeIPA nodes and perform the following set of steps:
 - a. Get the CA certificate from `/etc/jumpgate/config.toml` and grab the pinned CA certificate from the `agent.relayServerCertificate` parameter.
 - b. Configure your proxy server to trust this certificate for the CCM traffic.
 - c. Copy your proxy server's CA certificate and replace the contents of `agent.relayServerCertificate` in `/etc/jumpgate/config.toml`.
 - d. Configure your proxy to start MITM-ing the underlying TLS connection.